INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Prediction of COVID-19 trend in India using time series forecasting

Tabasum Mirza[1], Malik Mubasher Hassan[2]*, Mirza Waseem Hussain[1]

1 Lecturer Computer Science, Dept. of School Education Govt. of J&K, India
2 Assistant Professor, Dept of ITE, BGSB University Rajouri J&K, India. Tel.: +917298286036

OPEN ACCESS

Received: 22-07-2020
Accepted: 16-08-2020
Published: 01-09-2020

Editor: Dr. Natarajan Gajendran

Citation: Mirza T, Hassan MM, Hussain MW (2020) Prediction of COVID-19 trend in India using time series forecasting. Indian Journal of Science and Technology 13(32): 3248-3274. https://doi.org/10.17485/IJST/v13i32.1214

***Corresponding author**.

Tel: +917298286036
mmhassan@bgsbu.ac.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: COVID-19 pandemic is one of the prevalent challenges mankind has ever faced and there is a lot of uncertainty prevailing over the future with respect to COVID-19. In this situation machine learning algorithms can be useful for real-time analysis and prediction of trends of the infections. The objective of the research study is to analyze the COVID-19 trend in India and forecast the trend of outbreak in near future. This model can provisionally guide the government and healthcare organizations in making preparations for the upcoming situation arising out of COVID-19 transmission. **Methods:** The COVID-19 data from 30-Dec 2019 to 27-July 2020 was used for prediction of COVID-19 trend in next 30 days i.e. from 28 July to 26 August. The time series forecasting approaches with ARIMA Model and PROPHET were used for forecasting. The performance of these models was evaluated using validation metrics and good performance was indicated. **Findings**: The prediction results indicate an increasing trend of COVID-19 positive, active and deceased cases in India for next 30 days i.e. up to 26 August 2020. Novelty: COVID-19 pandemic is a new problem. The novelty and originality of this research lies in the fact that time series forecasting is used for real time analysis and prediction of COVID-19 pandemic.

Keywords: COVID-19; time series analysis; ARIMA model; time series forecasting; PROPHET

1 Introduction

Coronavirus disease (COVID-19) emerged as a mysterious viral respiratory disease in Wuhan, Hubei, China in Dec 2019 [(1)]. The cause of disease was identified as kind of corona virus also known as novel coronavirus with the formal name of SARS-CoV-2 by the International Committee on Taxonomy of Viruses, as it was showed similarity with SARS-2003.The virus can progress through the respiratory tract into a person's lungs causing inflammation and the air sacs, or alveoli, that can fill with

fluid and pus. This condition can limit a person's ability to take in oxygen and in severe cases; patients cannot take enough oxygen or breathe out carbon dioxide, finally leading to life threatening conditions like heart failure.

The corona virus outbreak came into limelight on 31st December 2019 when China informed WHO about the outbreak of pneumonia in Wuhan, Hubei province. COVID-19 emerged as a public health emergency very quickly and was declared as a global pandemic by WHO on 11 March 2020 [(2)]. As on 27th July 2020 World has unfortunately witnessed a total of 16.2 million COVID-19 infected cases with 0.648 million deaths and 940000 recoveries [(3)].

Most of the COVID infected cases reported symptoms like fever, dry cough, and tiredness with breathing difficulty in severe cases.COVID-19 infection can lead to potential complications like pneumonia, cytokine strokes and multi organ failures ultimately resulting in death of the patient [(4–6)]. Some of the cases are asymptotic without noticeable illness in the infected person but the asymptotic patients are also contagious virus carriers although their infectivity may be weak [(7,8)].

The mode of transmission of SARS-CoV-2 is through close direct or indirect contact with the infected persons via infected respiratory secretions or saliva. When an infected person coughs or sneezes, sings or talks within 1 meters of the range the virus can reach the mouth, eyes or nose of the susceptible person resulting in infection [(9)]. Contact with the surrounded objects or surfaces infected with virus can result in infection via indirect mode of transmission called fomite transmission [(10)]. Indoor settings with poor ventilation virus can spread through air as a result of airborne transmission. Incubation period for the virus is 14 days and in the majority of cases symptoms start to appear after 4-5 days of exposure while in some cases symptoms may appear as late as 11 days after exposure [(11)]. Scientific evidence suggests that 1-3 days before the development of symptoms is the peak time when an infected person can transmit the virus to another person [(12,13)]. Even though people from all age groups are at risk of infection, most of the infections were reported from the middle age group and the most vulnerable categories of population to virus includes senior citizens, pregnant women and persons with co-morbidities [(14)].

The COVID-19 patient can show the following symptoms:

- Rise in body temperature
- Tiredness
- Dry cough
- Shortness of breath
- Sputum production
- Loss of taste or smell
- Muscle pain and chills
- Nasal congestion or flu
- Vomiting or Diarrhea
- Soreness of throat
- Conjunctivitis
- Headache
- Skin rashes, or discoloration of fingers or toes

To curb transmission of infection, WHO suggests following guidelines:-

1. Quick identification, test and isolation of suspected cases with proper facilities
2. Identification of closed contacts of the infected person, test and quarantine of these people.

3. Use of fabric masks in public places to avoid community transmission.
4. Social distancing and avoiding crowded places.
5. Practicing frequent hand hygiene i.e. washing of hands with soap or use of alcohol based sanitizers.
6. Use of protective gears and medical masks by health care workers [13,14]

The first case of COVID-19 was reported in India on 30 January 2020 in Kerala when a student travelled from Wuhan to back his home in India [15].

To contain the virus transmission in the community, a number of countries imposed lockdown restrictions to ensure confinement of the general public in their homes. Indian Govt ordered nationwide lockdown for 21 days on 24 March 2020 which was further extended as a preventive measure against spread of COVID-19 infection when the number of reported COVID-19 cases came close to 500 restricting the movement of 1.32 billion populations. As on 27th July, 2020, India has reported 1.43 million COVID positive cases, 0.486 million active case, 32,771 deaths and 9.17 million recoveries. On 26th July 2020 India recorded highest number of positive cases numbering 50,000. Moreover, from the last five days the daily reported positive cases is more than 45,000 each day. The recovery rate from the disease is almost 64% in India and the case fatality report (CFR) is 2.3%. There are about 1300 testing laboratories in India and number is keeping on increasing. More than 0.5 million tests are conducted daily and there are 1.1 million isolation beds in the country and more than 11000 COVID care facilities. The CFR is progressively decreasing and it is a good sign. The lockdown declared by India 0n March 22, 2020 is continuing and it is the 125 day today.

Among Indian states and Union territories, Maharashtra reported the highest number of 3,75,799 cases with 13,756 deaths followed by Tamil Nadu having 2,13,723 cases and 3,494 deaths, Delhi is third worst affected city having 1,30,606 cases and 3827 fatalities [16,17].

As on date India stands at the third position in number of reported coronavirus cases, with the U.S. leading in number of 41,48,011 coronavirus cases and 148,012 deaths followed by Brazil having 2,419,091 cases and 87,004 deaths. National average of CFR (COVID-19 Fatality Rate) in India is 2.28% which is 1.98% less than the average global CFR and is lower than that of the United States (3.88 per cent) and Brazil (3.81), as on 20 July 2020 but it varies for the individual states with highest CFR of 4.09% in Gujarat, 3.55% in Maharashtra, and 2.57% in West Bengal [18,19].

As of now, there is no treatment for COVID-19 but symptoms may be treated with a certain combination of drugs or plasma transfusion for severe patients depending on clinical conditions of the patient treated [20,21]. Even though a number of countries like Russia, India, U. S have started human trials on vaccines to protect from the disease but there is no reported success till date [22–24].

Time Series forecasting can be used to predict the number of COVID-19 cases, deaths and recoveries in the near future. We are using time series forecasting approaches with ARIMA model and PROPHET to predict the COVID-19 cases in India for next 30 days. Although it is novel coronavirus and the data available at the initial stage was very small, therefore there are chances of some uncertainty or inaccuracy with forecasted trend of infections [25] [26]. ARIMA Models have already been used for the prediction of infectious diseases or other natural calamities and is suitable for short term predictions based on historical data [27–33]. Number of studies have been conducted to predict future trends of COVID-19 using various statistical models but there are few limitations like lack of proper data, unreported cases, lack of testing, over fitting of data, use of improper model and dynamically changing situations leading to unpredictable increase or decrease in the number of cases [34–39].

Keeping in view the increasing trend of the COVID-19 and subsequent rise of fear among the people motivated us to carry on the research to forecast the cases in the coming days, so that the necessary preparedness can be executed [40]. The paper is organized into five sections. Section 1 gives the overall introduction of the work carried out. Section 2 depicts the objectives clearly. Section 3 describes the methodology adopted along with the detailed flow of procedures. Section 4 discusses the results and findings and in the last section 5 gives the summary of the work done.

2 Objectives

1. To predict number of COVID-19 cases,
2. To predict number of recovered cases and
3. To predict number of deceased cases in India from 27-July to 26 August.

3 Methodology

The flowchart of the methodology adopted here in this study is shown in Figure 1.

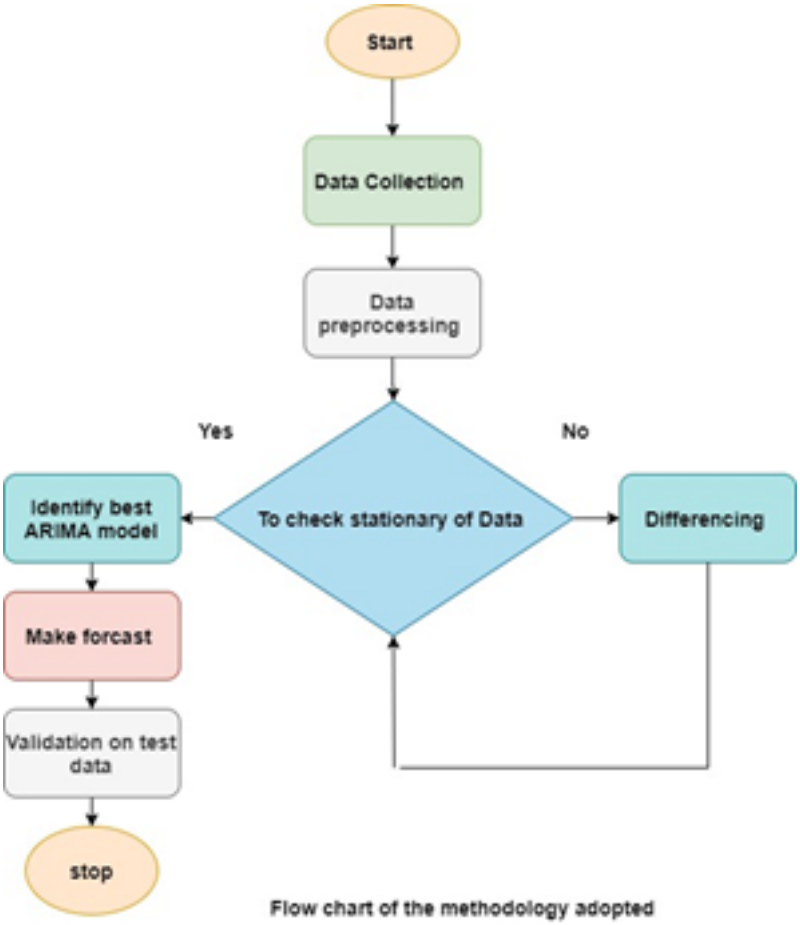


Fig 1. Methodology adopted

1. Data Collection. COVID-19 data is easily available for research and analysis purposes. Global COVID-19 data was collected from the website of WHO COVID-19 situation report and the Indian COVID-19 data was retrieved from the official Indian website covid19india.org andworldometers.info. We have collected the COVID-19 report data from a comma-separated file (CSV), and uploaded that in Excel 2013.Then data for J&K was extracted and converted into Time series data to be used for analysis and forecasting in R studio software.
2. Analysis of Data. Data was analyzed for missing values and then data was converted to time series data for use with the ARIMA Model and PROPHET to predict COVID-19 trend in J&K and India for active cases, confirmed cases, recoveries and deceased cases.
3. Forecasting. Time series forecasting with ARIMA Model and PROPHET is used for forecasting the trend of infections in near future.
4. Validation. Models are validated on the basis of comparison of actual test data with the forecasted data by dividing the dataset into training set and testing set. The model prediction is then tested on test data of 20 days with actual data to check the predictive capability of the model.

3.1. ARIMA

It is an acronym for 'Auto Regressive Integrated Moving Average' i.e. a class of statistical models that can be used for modeling stationary time series data on the basis of its past values, lags between observations and the lagged forecast errors, used to forecast or predict future values. The ARIMA model has three components:

- AR(Auto Regression)
- Integrated- representing the degree of differencing to make time series stationary
- MA(Moving Average)

The ARIMA model is denoted as ARIMA (n, f, p), in which n determines the order of auto regression terms, f is the degree of differencing used to form a stationary times series, and p is the order of moving average. The value of 'v' at time t for an ARIMA model is estimated as equation (1). Here the moving average parameter is denoted by $'\varnothing'$

$$\text{At f} = 0; \quad v_t = V_t$$

$$\text{At } f = 1;\ v_t = V_t - V_{t-1}$$

$$\text{At f} = 2; (v_t = V_t - V_{t-1}) - (v_t = V_t - V_{t-2}) = V_t - 2V_{t-1} + 2V_{t-2}$$

At f=2, the second difference is actually the 1st difference of the 1st difference and not the difference of the 2 periods ago. It is not the local trend but, it is the local acceleration of the series.

The general forecasting equation in terms of v is given by:

$$v_t = \mu + \varnothing_1 V_{t-1} + \varnothing_n V_{t-n} - \propto_1 e_{t-1} - \cdots - \propto_p e_{t-p} \tag{1}$$

The time series is checked for stationary using Augmented Dickey Fuller (ADF) test and may require logging or differencing of terms to make it stationary and to stabilize the series[41].

3.2 PROPHET

It is an open source Face book library that can be used for decomposition of time series and forecasting of future trends easily and accurately. This model is flexible in nature and can deal well with missing values. The time series model is additive in nature and fits linear data. It has three main components:

- Trend
- Seasonality
- Holidays

The general equation for PROPHET model is given as:

$$y(t) = g(t) + s(t) + h(t) + e(t) \tag{2}$$

Where g(t) logistic growth curve for modeling non periodic changes in time series. s(t) is periodic changes, h(t) is effectiveness of yearly seasonally holidays and e(t) is error term accounts unusual changes not accommodated by model.

4 Results and Findings

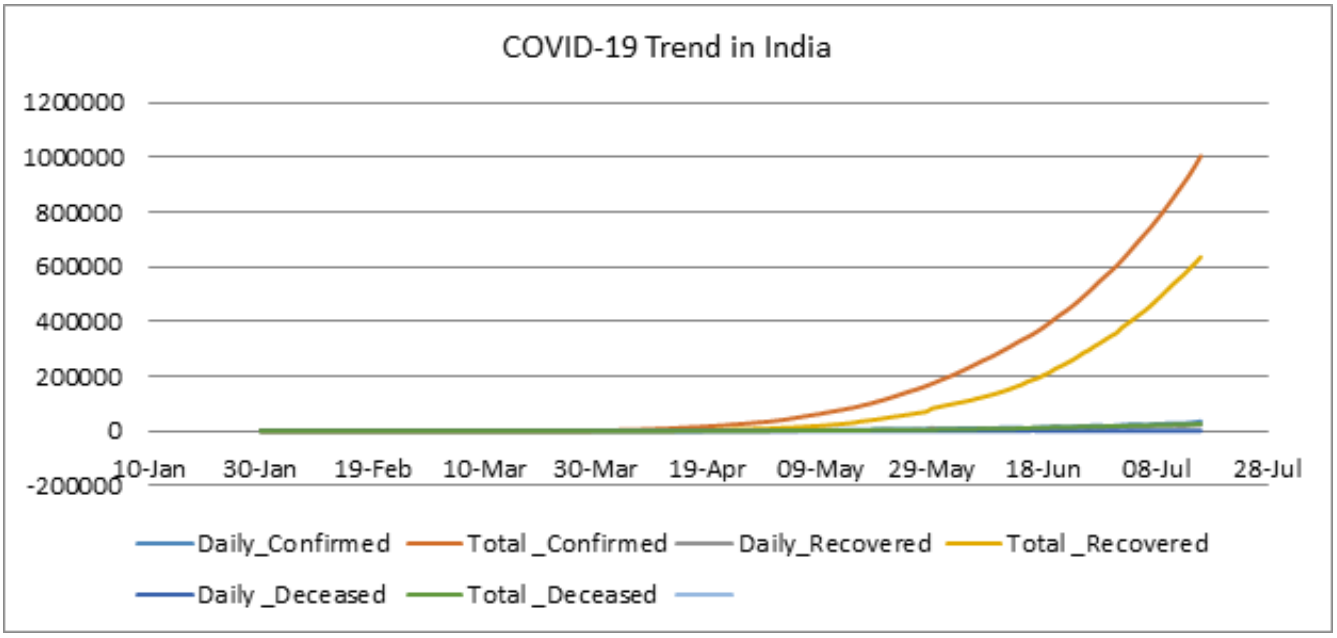


Fig 2. Analysis of COVID-19 trend in India

Figure 2 displays the COVID-19 trend for India for the daily confirmed, total confirmed, daily recovered, total recovered, daily deceased and total deceased cases from 30 Jan 2020 to 20-July 2020.The data was plotted to visualize the count of confirmed cases, deceased cases and recovered cases in India. Graph indicates continuously ascending trend in all components. The visualization indicates that number of confirmed COVID-19 cases rose from 1 on 30 January, 2020 to 1.1 million cases on 20 July, 2020.X-axis in the plot denotes dates and Y-axis denotes the number of cases.

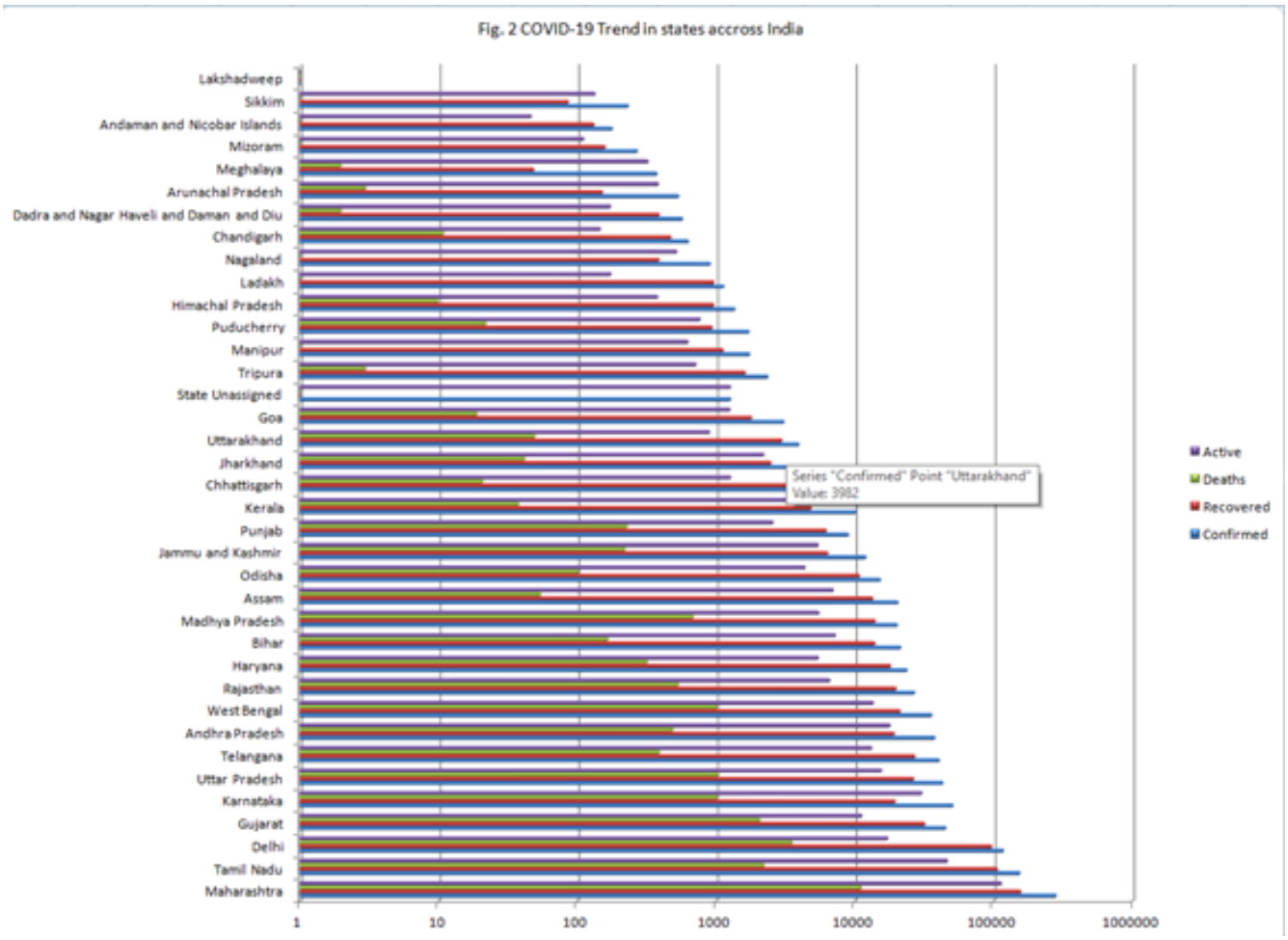


Fig 3. State wise COVID-19 trend for active, confirmed, recovered and deceased cases.

The Figure 3 displays Indian state wise COVID-19 trend for active, confirmed, recovered and deceased cases. The total cases are increasing with time but certain states are more affected like Maharashtra, Tamil Nadu, Delhi and Gujarat as indicated in the figure.

In this research study, time series analysis and forecasting using Arima model and Prophet to illustrate the trend of infections from the period of 30 January to 27 July and to forecast the future cases from 28 July to 26 August.

4.1 Forecast using ARIMA model

The visualization displays residual plots for confirmed, deceased, recovered and active COVID-19 cases in the country from 30 Jan 2020 to 27 July 2020.Time series is non stationary and non seasonal in nature. Increasing trend indicates the continuous surge in number of COVID-19 cases. Daily data of covid-19 cases acts as a variable for time series data model against time. We are using Auto ARIMA for the order of auto regression to find the value of order variables.

Figures 4, 5 and 6 indicates ACF- the coefficients of correlation indicating relationship between a time series and its lags and PACF - the partial correlation coefficients for ARIMA models.ACF and PACF are used to indicate the relationship between the observations in time series and its lags for the time series and

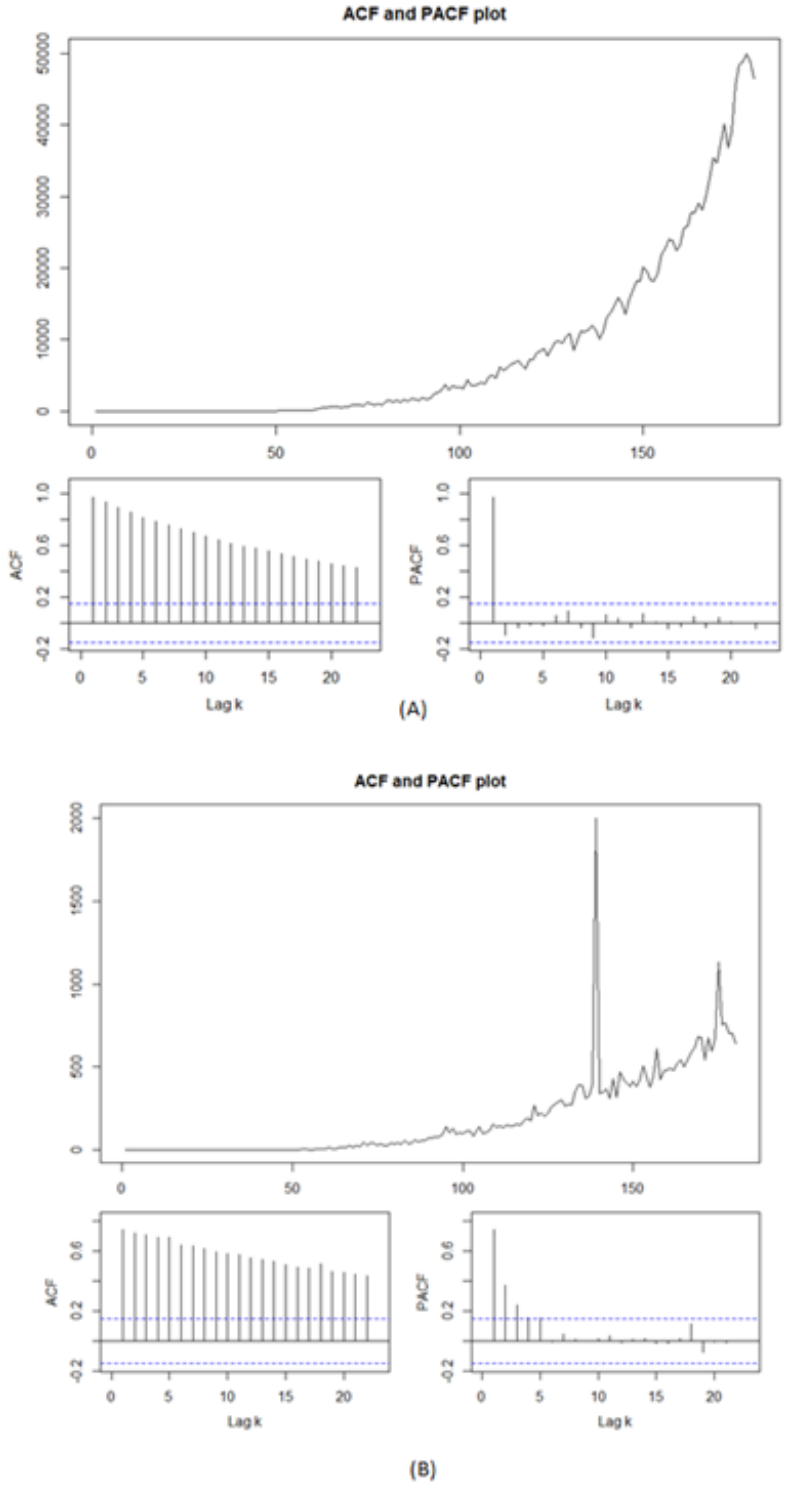
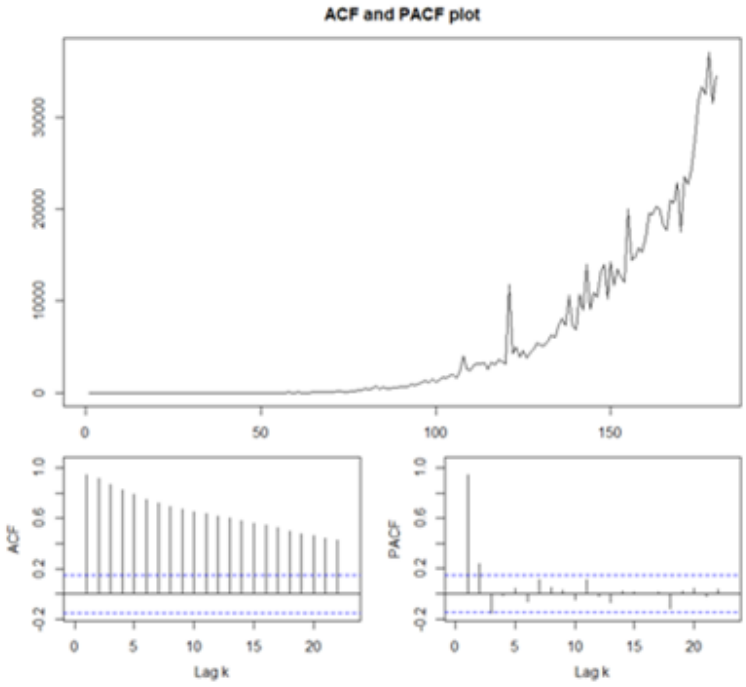
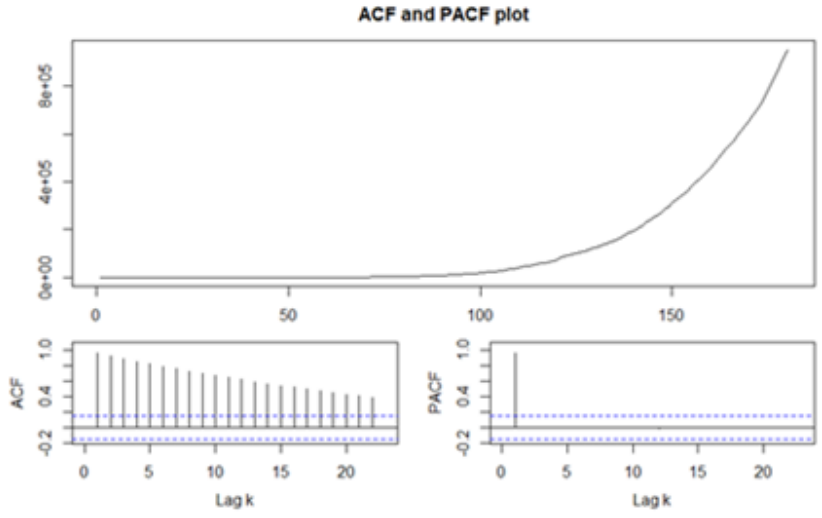


Fig 4. (A)and (B). ACF and PACF plots for daily confirmed cases and daily deceased cases respectively

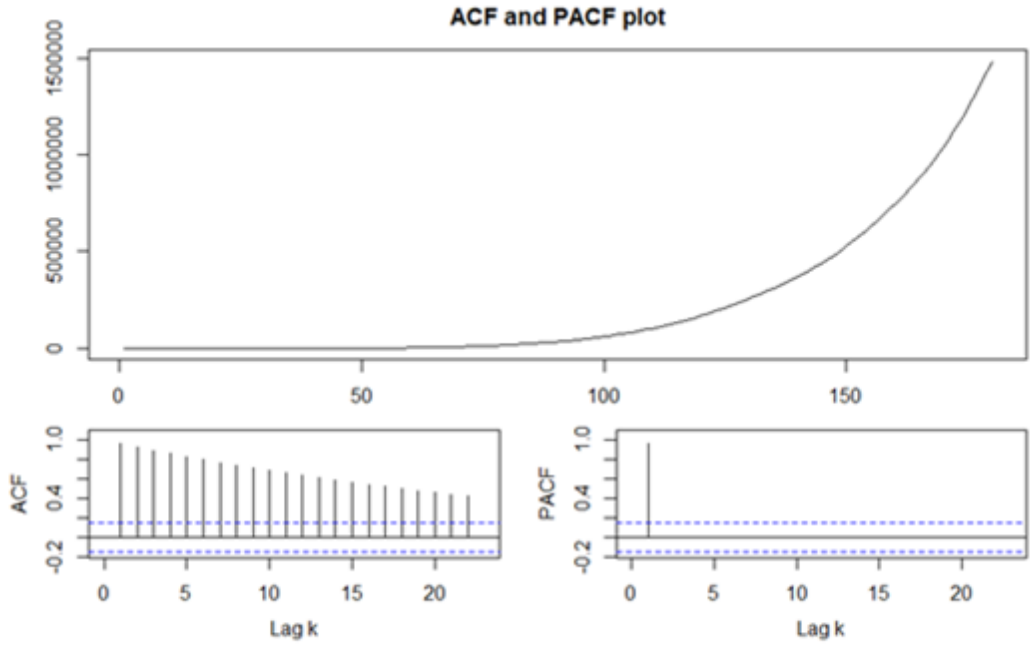


(c)

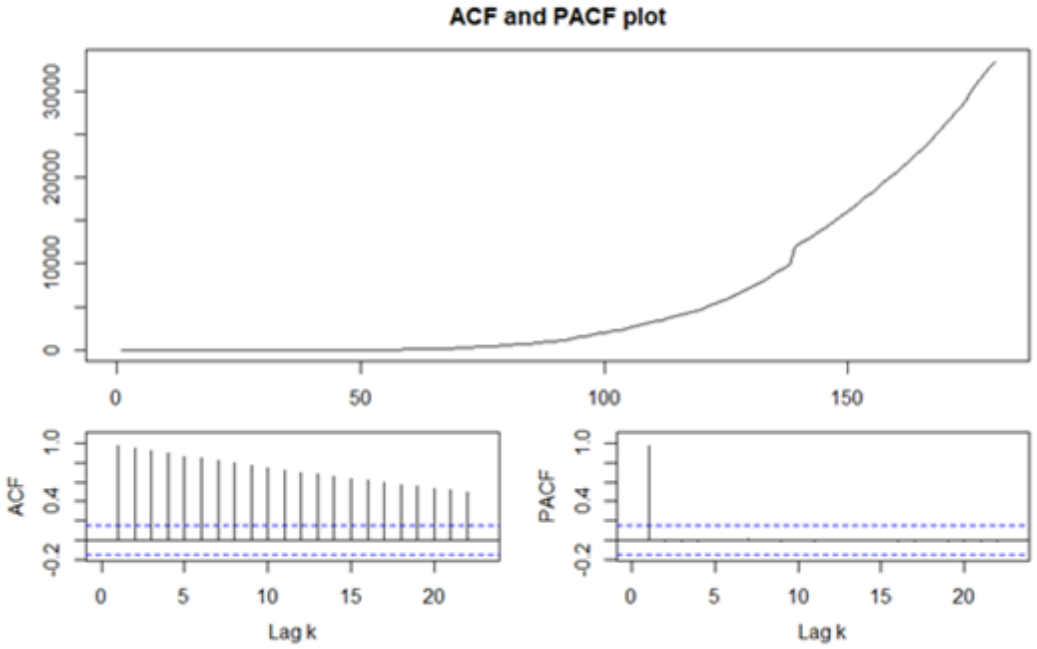


(D)

Fig 5. (C) and (D). ACF and PACF plots for daily recovered cases and total confirmed cases respectively



(E)



(F)

Fig 6. (E) and (F). ACF and PACF plots for total recovered cases and total deceased cases respectively

for determining the order parameters for ARIMA model. The plots show that the ACF and PACF values of the time series are declining and almost close to zero.

Figure 7 shows error Residuals vs. fitted values. The dotted line at y=0 indicates our fit line. Any point on the fit line has zero residual. Points above have positive residuals and points below have negative residuals. The red line is the smoothed high order polynomial curve to give us an idea of the pattern of residual movement. In our case we can see that our residuals have slightly curved patterns showing slight deviation in results indicating the presence of outliers in data and some errors but the assumption of normal distribution holds true.

The figure also displays Normal Q-Q plot used to check whether our residuals are following Normal distribution. It displays scale location plot to indicate the spread of points across predicted values range to assume homoscedasticity in regression. Figure also indicates Cook's distance to find influential outliers in the predictor values set and illustrates the histogram of residuals.

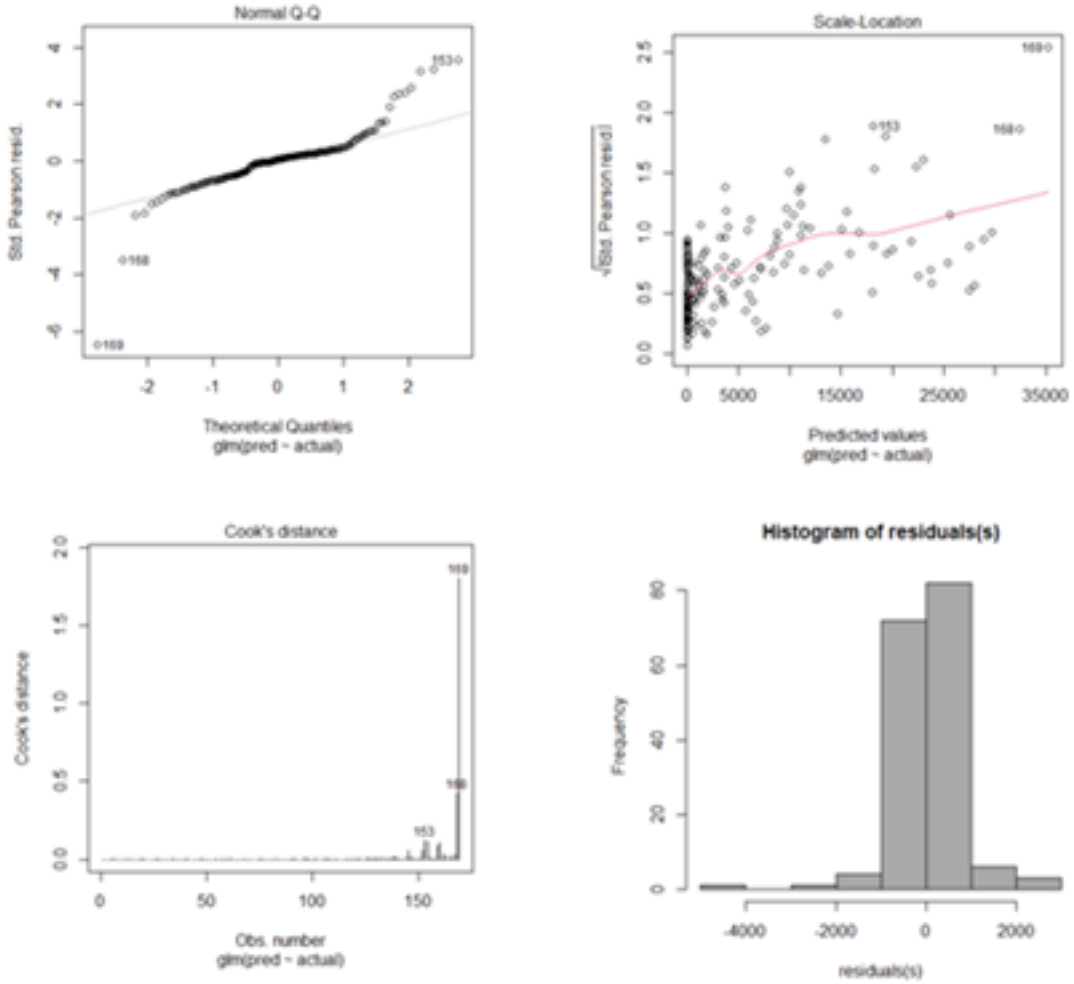


Fig 7. Residual plot

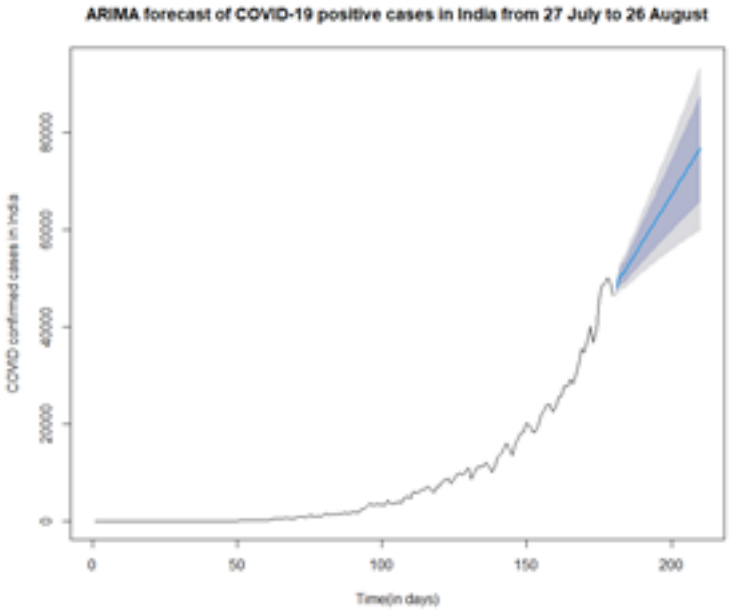


Fig 8. Forecast plot for ARIMA (2, 2, 1)

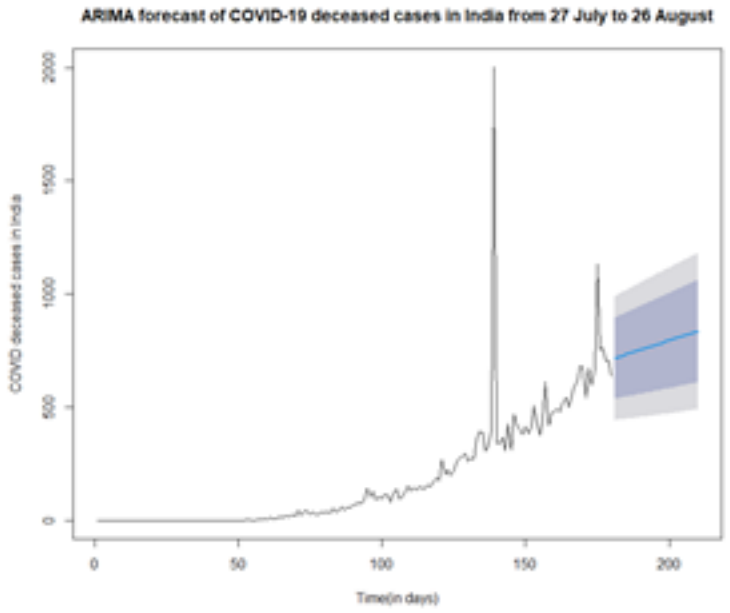


Fig 9. Forecast plot for ARIMA (0, 1, 1)

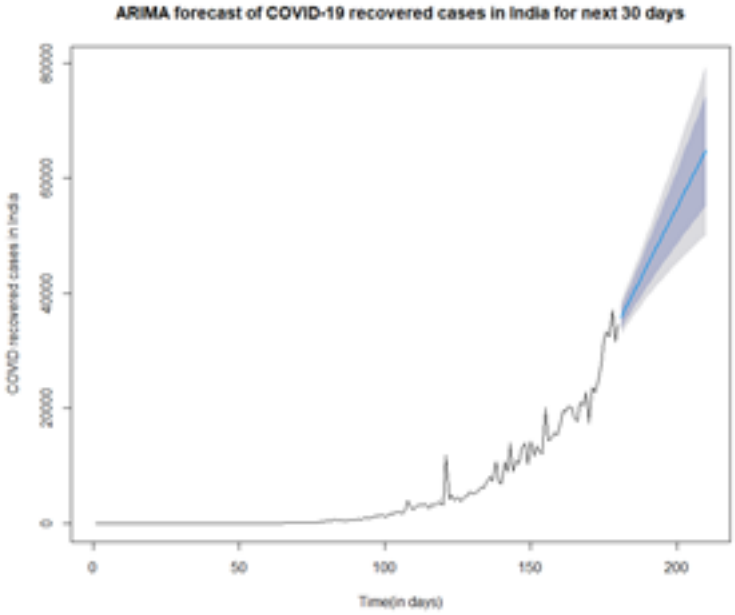


Fig 10. Forecast plot for forecast plot for ARIMA (0, 2, 2)

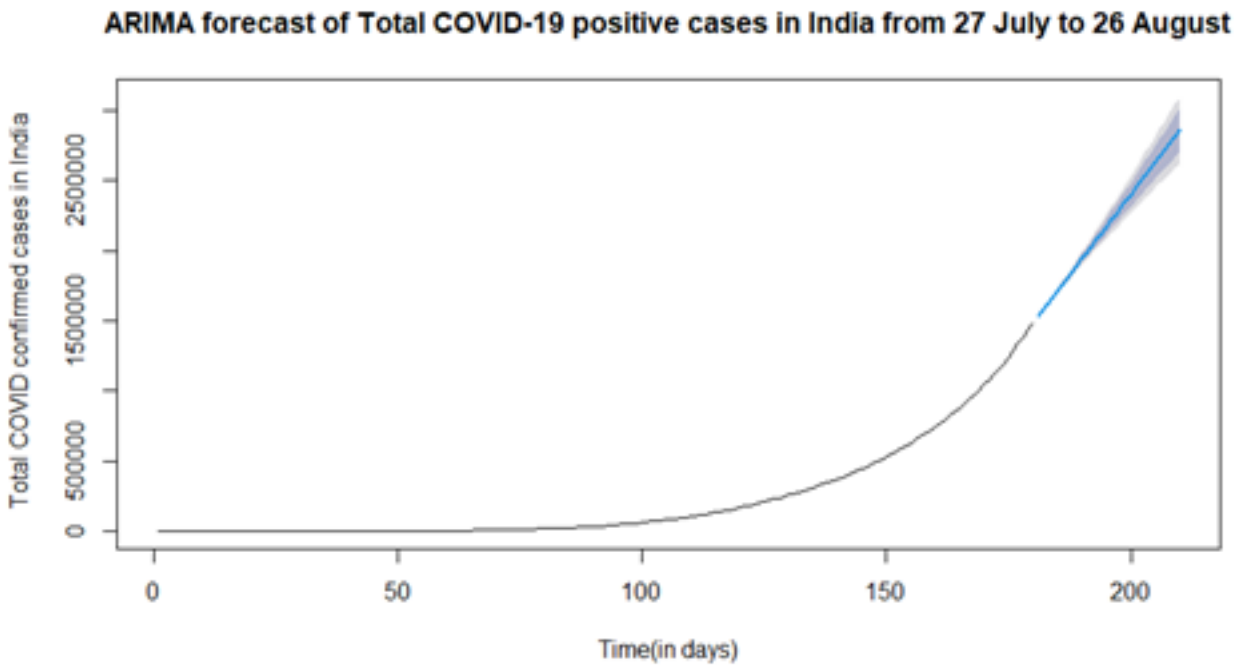


Fig 11. Forecast plot for ARIMA (0, 2, 1)

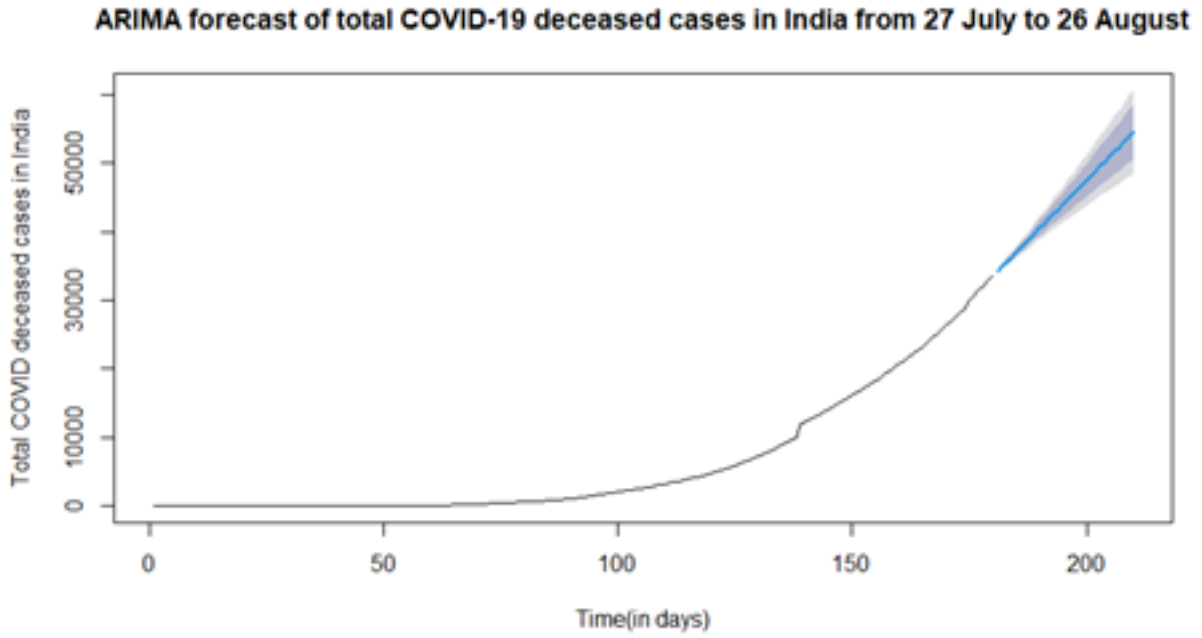


Fig 12. Forecast plot for ARIMA (0, 2,1)

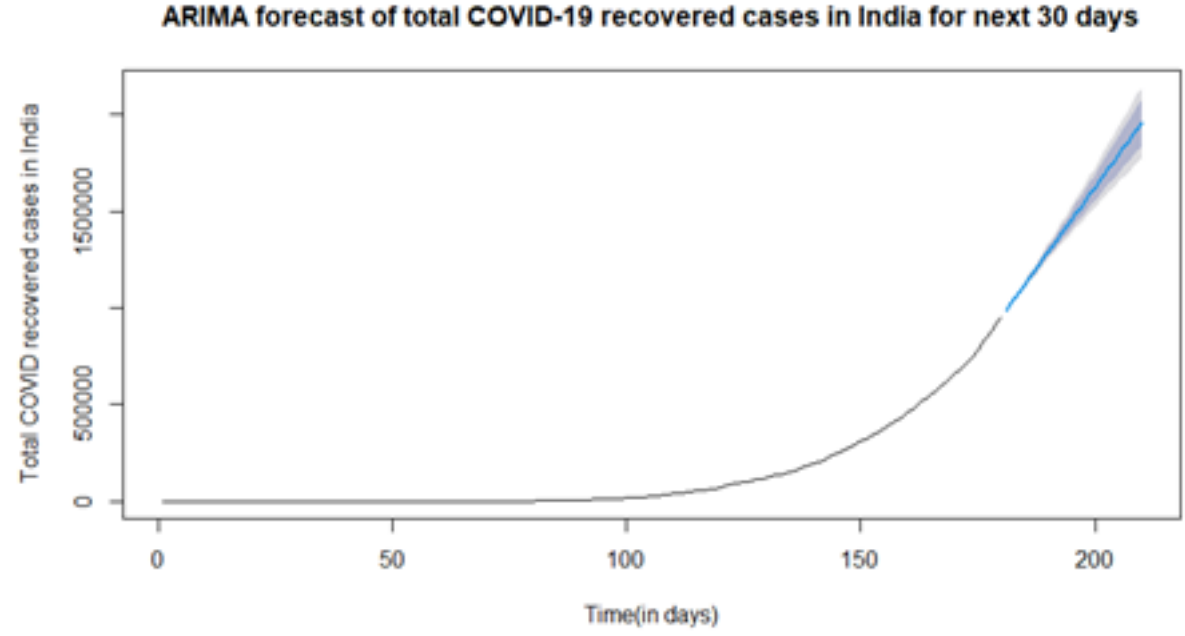


Fig 13. Forecast plot for ARIMA (2, 2,1)

Figure 8 displays forecasted COVID-19 trend in India for positive cases using ARIMA Model of COVID-19 confirmed positive cases in India from 30 January 2020-26 August 2020. The visualization indicates continuous increase in the number of positive cases with time. Trend for the number of daily recovered and daily deceased cases is indicated in Figures 9 and 10 respectively. Trend for the number of total positive, deceased, and cases in India from 30th January to August 26, 2020 is shown in the Figures 11, 12 and 13 respectively.

Tables 1, 2 and 3 show the forecasted daily confirmed, recovered and deceased cases from July 28 to August 26, 2020 respectively with 95% confidence interval. Tables 4, 5 and 6 show the forecasted total confirmed, recovered and deceased cases from July 28 to August 26, 2020 respectively with 95% confidence interval.

Table 1. Forcasted daily confirmed cases from July 28-August 26, 2020 with 95% CI

Day	Forecast	Lo 95	Hi 95
181	48022	46188	49856
182	50047	47330	52765
183	50865	47722	54008
184	51507	47957	55056
185	52514	48483	56545
186	53583	49070	58095
187	54542	49572	59511
188	55479	50054	60904
189	56450	50559	62340
190	57427	51067	63788
191	58395	51563	65228
192	59361	52051	66670
193	60329	52536	68122
194	61298	53017	69580
195	62267	53490	71043
196	63235	53957	72512
197	64203	54418	73988
198	65172	54873	75470
199	66140	55321	76958
200	67108	55763	78453
201	68077	56199	79955
202	69045	56627	81463
203	70013	57050	82977
204	70982	57466	84498
205	71950	57875	86025
206	72918	58278	87558
207	73887	58675	89098
208	74855	59065	90645
209	75823	59449	92198
210	76792	59827	93757

Table 1.

Table 2. Forcasted daily recovered cases from July 28-August 26, 2020 with 95% CI

Day	Forecast	Lo 95	Hi 95
181	35877	32976	38779
182	36877	33753	40001
183	37877	34510	41244
184	38877	35246	42508
185	39877	35964	43790
186	40877	36664	45090
187	41877	37348	46405
188	42877	38018	47736
189	43877	38673	49081
190	44877	39314	50439
191	45877	39944	51810
192	46876	40561	53192
193	47876	41166	54587
194	48876	41761	55992
195	49876	42345	57408
196	50876	42919	58834
197	51876	43482	60270
198	52876	44037	61715
199	53876	44582	63170
200	54876	45118	64634
201	55876	45645	66106
202	56876	46164	67587
203	57876	46674	69077
204	58876	47176	70575
205	59875	47671	72080
206	60875	48157	73594
207	61875	48635	75115
208	62875	49106	76644
209	63875	49570	78181
210	64875	50026	79724

Table 2.

Table 3. Forcasted daily deceased cases from July 28-August 26, 2020 with 95% CI

Day	Forecast	Lo 95	Hi 95
181	717	443	991
182	721	445	997
183	725	446	1004
184	729	447	1011
185	733	449	1018
186	737	450	1025
187	742	452	1032
188	746	453	1038
189	750	454	1045
190	754	456	1052
191	758	458	1059
192	762	459	1065
193	766	461	1072
194	770	462	1078
195	775	464	1085
196	779	466	1092
197	783	467	1098
198	787	469	1105
199	791	471	1111
200	795	472	1118
201	799	474	1124
202	803	476	1131
203	807	478	1137
204	812	479	1144
205	816	481	1150
206	820	483	1157
207	824	485	1163
208	828	487	1169
209	832	489	1176
210	836	491	1182

Table 3.

Table 4. Forcasted total confirmed cases from July 28-August 26, 2020 with 95% CI

Day	Forecast	Lo 95	Hi 95
181	1528421	1526458	1530385
182	1574339	1569455	1579223
183	1620256	1611737	1628776
184	1666174	1653421	1678926
185	1712091	1694582	1729600
186	1758008	1735272	1780745
187	1803926	1775528	1832324
188	1849843	1815382	1884304
189	1895761	1854859	1936662
190	1941678	1893980	1989376
191	1987596	1932763	2042428
192	2033513	1971222	2095804
193	2079430	2009371	2149490
194	2125348	2047222	2203473
195	2171265	2084786	2257744
196	2217183	2122072	2312293
197	2263100	2159090	2367110
198	2309017	2195845	2422189
199	2354935	2232347	2477522
200	2400852	2268602	2533103
201	2446770	2304615	2588924
202	2492687	2340393	2644981
203	2538604	2375941	2701268
204	2584522	2411264	2757779
205	2630439	2446367	2814512
206	2676357	2481253	2871460
207	2722274	2515928	2928620
208	2768192	2550395	2985988
209	2814109	2584657	3043561
210	2860026	2618719	3101334

Table 4.

Table 5. Forcasted total recovered cases from July 28-August 26, 2020 with 95% CI

Day	Forecast	Lo 95	Hi 95
181	986467	983480	989455
182	1019789	1014524	1025055
183	1053736	1045480	1061991
184	1086870	1075154	1098587
185	1120776	1105363	1136190
186	1154061	1134470	1173651
187	1187786	1163827	1211746
188	1221237	1192533	1249941
189	1254831	1221164	1288498
190	1288373	1249453	1327293
191	1321912	1277507	1366317
192	1355482	1305352	1405612
193	1389011	1332928	1445095
194	1422579	1360328	1484830
195	1456116	1387482	1524750
196	1489675	1414458	1564892
197	1523220	1441219	1605221
198	1556773	1467797	1645748
199	1590323	1494184	1686461
200	1623872	1556208	1727355
201	1657423	1584841	1768429
202	1690973	1613358	1809674
203	1724524	1641765	1851092
204	1758073	1670062	1892674
205	1791624	1648827	1934421
206	1825173	1674021	1976327
207	1858724	1754324	2018390
208	1892274	1782208	2060606
209	1925824	1809992	2102975
210	1959375	1837679	2145492

Table 5.

Table 6. Forcasted total deceased cases from July 28-August 26, 2020 with 95% CI

Day	Forecast	Lo 95	Hi 95
181	34170	33892	34448
182	34873	34441	35304
183	35576	35000	36152
184	36279	35557	37000
185	36982	36112	37852
186	37685	36662	38708
187	38388	37207	39569
188	39091	37746	40435
189	39794	38281	41307
190	40496	38809	42184
191	41199	39333	43066
192	41902	39851	43953
193	42605	40365	44846
194	43308	40873	45744
195	44011	41376	46647
196	44714	41874	47554
197	45417	42368	48466
198	46120	42857	49384
199	46823	43341	50305
200	47526	43821	51231
201	48229	44296	52162
202	48932	44767	53097
203	49635	45233	54036
204	50338	45696	54980
205	51041	46154	55927
206	51744	46608	56879
207	52447	47059	57835
208	53150	47505	58794
209	53853	47947	59758
210	54555	48386	60725

Table 6.

The ARIMA model predicts that the daily number of positive cases will increase in coming days can reach up to 93756 cases in the worst case scenario and 59826 in best case scenario. Average count of daily forecasted cases on 26 August is 76791 with 95% confidence Interval (CI). The count of daily recoveries will also increase and can reach up to lower limit of 50025, average count of 64875 and upper limit of 79724 on 26 August. The count of daily deceased cases will increase and can reach up to lower limit of 490, average count of 836 cases and upper limit of 1182 on 26 August.

The ARIMA model predicts that the total number of positive cases will increase in coming days can reach up to 3101334 cases in the worst case scenario and 2618719 in best case scenario. Average total count of forecasted cases on 26 August is 2860026 on 95% confidence Interval (CI).

It is observed that the number of deaths as well as recoveries increase with due course of time. The recovery rate of India is increasing day by day and is higher than CFR. The count of total recoveries will also increase and can reach up to lower limit of 1773257, average count of 1959374 and upper limit of 2145491 on 26 August.

The count of total deceased cases will increase and can reach up to lower limit of 48385, average count of 54555 cases and upper limit of 60725 on 26 August. [26–42]

Validation of models

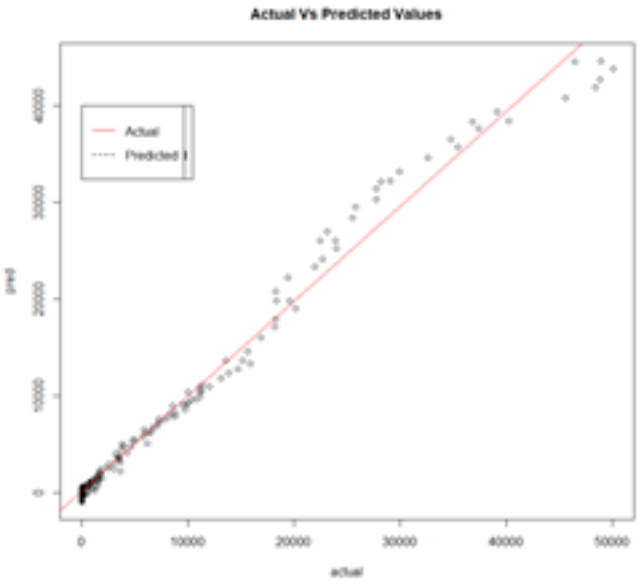


Fig 14. Actual vs Predicted Values for Arima model

The visualization plot represented by Figure 14 is used for comparison of actual and forecasted cases by ARIMA model to verify the efficiency of the models used for forecasting.

The data was divided into training data set from 30 January to 27 July and test data set for verification from 28 July to 26 August. Then, the actual cases were plotted against the forecasted cases from 7 July to 27 July to visualize the precision of forecasted values.

Table 7 indicates goodness of fit of the models used for forecasting. In this study, we used three performance measures, namely ME, MAE and RMSE. The low values of MAE, RMSE and ME indicate good fit of the models. The MAE is smaller than the RMSE and gives indication of low errors in the variance.

Table 7. Parameters of best fit ARIMA models

Model	ARIMA order	ME	RMSE	MAE
Daily positive cases in India	Arima(2,2,1)	96.60603	922.6173	522.4107
Daily deceased cases in India	Arima(0,1,1)	0.6265458	138.4616	44.19856
Daily recovered cases in India	Arima(0,2,2)	159.7596	1463.972	666.9492
Total positive cases in India	Arima(0,2,1)	199.6946	993.4804	556.431
Total deceased cases in India	Arima(0,2,1)	20.85905	140.6243	38.28361
Total recovered cases in India	Arima(2,2,1)	116.4499	502.871	198.0169

RMSE-Root Mean Square Error, **ME**-Mean Error, **MAE**- Mean Absolute Error.

4.2 Forecast using PROPHET

Time series forecasting using PROPHET is used to plot the observed trend of infections. Figures 15 and 16 illustrates the trend of COVID-19 infections for 180 days i.e. from the 30th January to 27th July in India.

Figures 17, 18 and 19 show the forecast for daily confirmed, recovered, and deceased cases in India respectively from 28 July to 26 August using PROPHET. Figures 20, 21 and 22 show the forecast for total confirmed, recovered, and deceased cases in India respectively from 28 July to 26 August using PROPHET.

The PROPHET time series forecasting predicts that the count of average daily positive cases can reach up to the count of 71727 cases on Aug 26 with count of 45779 daily recovered cases and 927 daily deceased cases. The forecast illustrates that total COVID-19 confirmed cases will rise to the count of 2007084, total recovered will reach up to 1316682, and total count of deceased cases in the country will reach up to 45679 in average scenario by 26 August 2020. Figure 23 illustrates Actual VS Predicted Values for PROPHET to indicate the accuracy of forecasting.

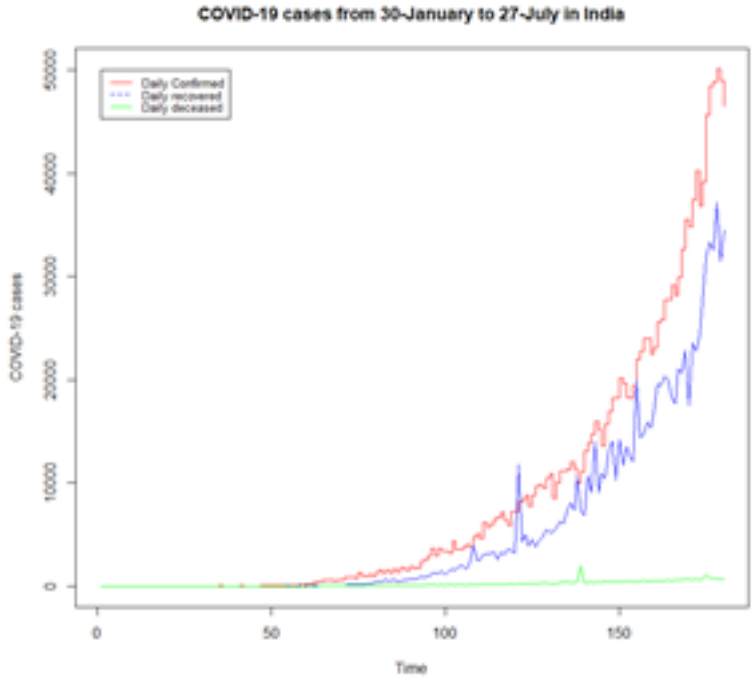


Fig 15. Cases from Jan-3- toJul 27

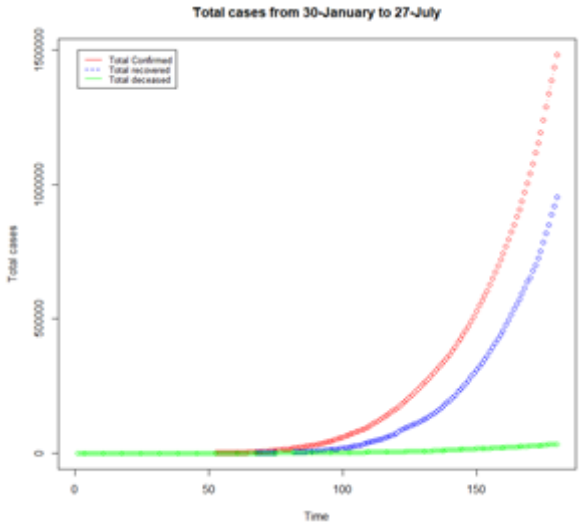


Fig 16. Total cases from Jan-3- to Jul 27

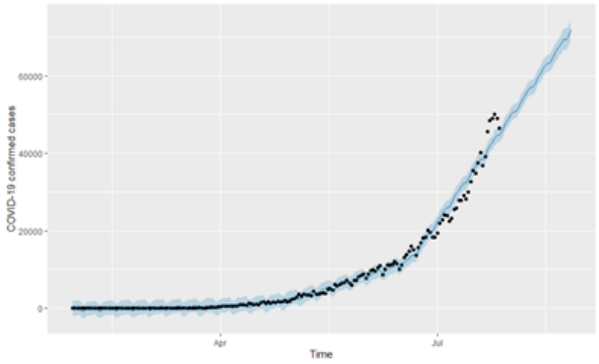


Fig 17. Forecast for daily confirmed cases in India from 28 July to 26 August using PROPHET.

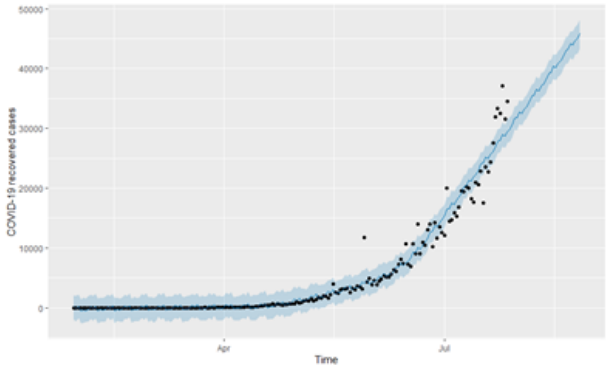


Fig 18. Forecast for daily recovered cases in India from 28 July to 26 August using PROPHET.

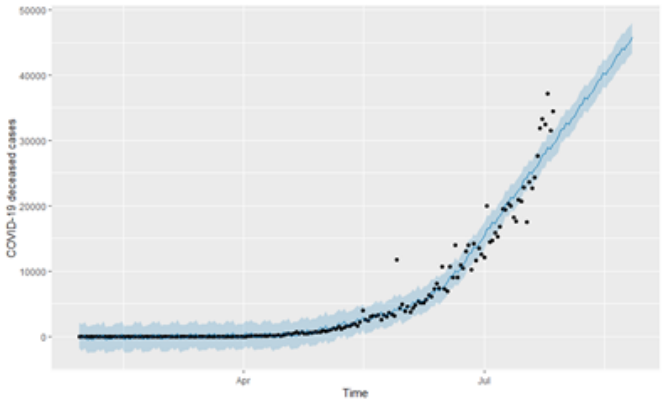


Fig 19. Forecast for daily deceased cases in India from 28 July to 26 August using PROPHET.

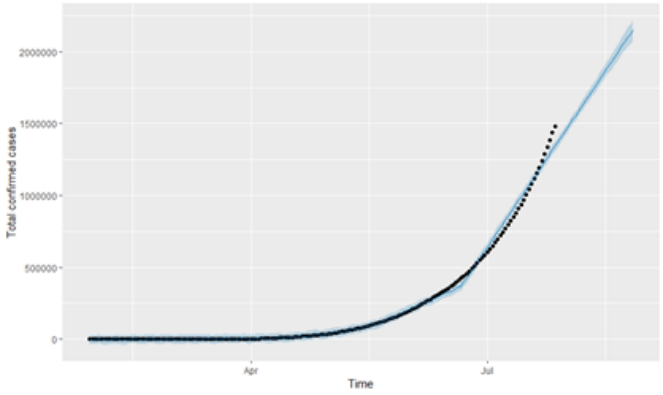


Fig 20. Forecast for total confirmed cases in India from 28 July to 26 August using PROPHET.

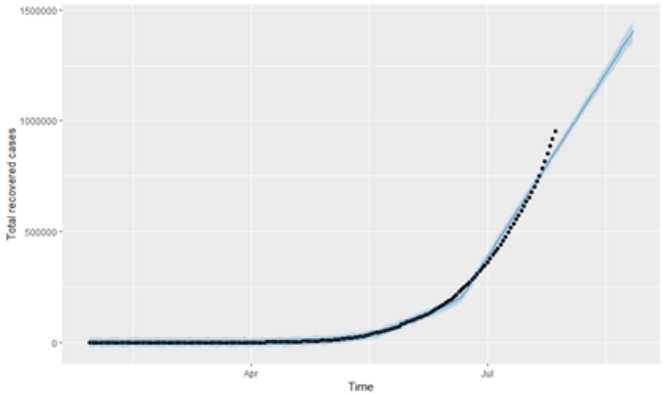


Fig 21. Forecast for total recovered cases in India from 28 July to 26 August using PROPHET.

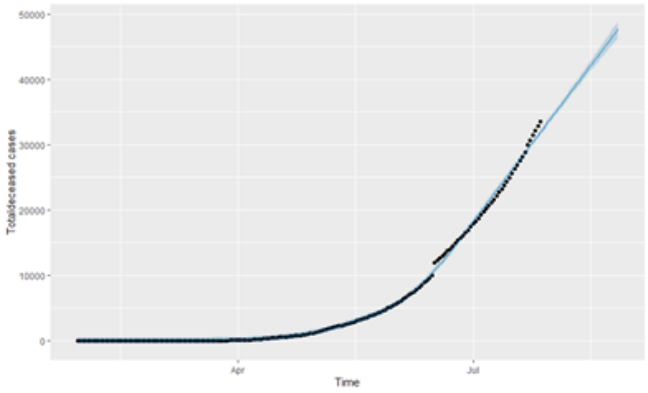


Fig 22. Forecast for total deceased cases in India from 28 July to 26 August using PROPHET.

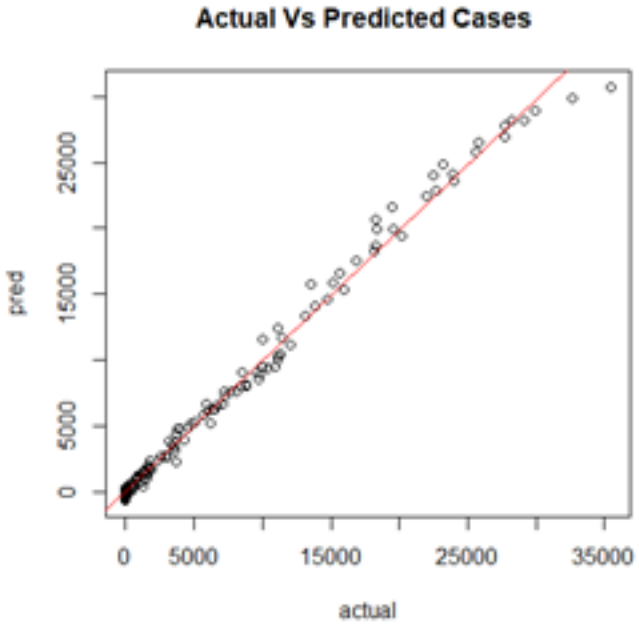


Fig 23.

5 Conclusion

Although the information about the Novel COVID-19 virus is evolving dynamically and not much is known about its behavior, mathematical models and machine learning algorithms can be used to predict the trend of active, positive, recovery and death cases. This research study helps in the real time analysis and forecasting of COVID-19 trend in India. Time series forecasting using ARIMA model and PROPHET were used for prediction of daily and total positive, recovered and deceased COVID-19 cases in India from 28 July to 26 August and results are indicating an increasing upward trend in the forecasted time of 30 days. This means that India has entered the stage of community transmission of SARS-Cov 2 virus. The number of daily deceased cases is much lower than the recovered cases that indicate low possibilities of causalities from the infections. India will have to ensure social distancing and other safety precautions to contain the spread of virus. We validated precision and accuracy of models using RMSE, ME and MAE, and

the validation results showed good regression fit and accuracy in prediction indicating good forecasting performance. As the situation with regard to present COVID-19 pandemic is keep on changing, therefore the exact prediction seems to be little difficult.

Acknowledgment

We express our deep sense of gratitude to the BGSBU for providing assistance under TEQIP-III grant.

References

1) Cheng CCV, Lau KPS, Woo CYP, Yuen KY. Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clinical Microbiology Reviews*. 2007;20:660–694. Available from: https://dx.doi.org/10.1128/cmr.00023-07.
2) WHO. Coronavirus (COVID-19) events as they happen. 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.
3) Worldometer. Coronavirus Cases. 2020. Available from: https://www.worldometers.info/coronavirus.
4) Divani AA. Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information. 2020.
5) Fan H, Tang X, Song Y, Liu P, Chen Y. Influence of COVID-19 on Cerebrovascular Disease and its Possible Mechanism. *Neuropsychiatric Disease and Treatment*. 2020;16:1359–1367. Available from: https://doi.org/10.2147/NDT.S251173.
6) Zhang P. The novel coronavirus (COVID-19) pneumonia with negative detection of viral ribonucleic acid from nasopharyngeal swabs: a case report. 2020.
7) Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. *Health Psychology*. 2020;39(5):355–357. Available from: https://dx.doi.org/10.1037/hea0000875.
8) Gao M, Yang L, Chen X, Deng Y, Yang S, Xu H. 2020.
9) Santarpia JL. Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center. 2020. Available from: https://doi.org/10.1101/2020.03.23.20039446.
10) Zhou J. Investigating SARS-CoV-2 surface and air contamination in an acute healthcare setting during the peak of the COVID-19 pandemic in London. 2020. Available from: https://doi.org/10.1101/2020.05.24.20110346.
11) Luo L. Modes of contact and risk of transmission in COVID-19 among close contacts. Cold Spring Harbor Laboratory Press. 20042606. Available from: https://doi.org/10.1101/2020.03.24.20042606.
12) Ma J. Exhaled breath is a significant source of SARS-CoV-2 emission," medRxiv. 2020. Available from: https://doi.org/10.1101/2020.05.31.20115154.
13) Du J. Psychological symptoms among frontline healthcare workers during COVID-19 outbreak in Wuhan. *Gen Hosp Psychiatry*. 2020. Available from: https://doi.org/10.1016/j.genhosppsych.2020.03.011.
14) Since. "Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information ,". 2020.
15) D. Corporation, "NRDC.". .
16) Jamwal A, Bhatnagar S, Sharma P. Coronavirus disease 2019 (COVID-19): Current literature and status in India. 2020. Available from: https://doi.org/10.20944/preprints202004.0189.v1.
17) WHO, COVID19 STATEWISE STATUS. 2020. Available from: https://www.mygov.in/corona-data/covid19-statewise-status/.
18) Info A. Outbreak prediction of covid-19 in most susceptible countries. 2021. Available from: https://doi.org/10.22034/GJESM.2019.06.SI.02.
19) "coronavirus Update. "Coronavirus Update (Live): 14,664,059 Cases and 609,279 Deaths from COVID-19 Virus Pandemic - Worldometer.". 2020. Available from: https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?%22.
20) Rajendran K, Krishnasamy N, Rangarajan J, Rathinam J, Natarajan M, Ramachandran A. Convalescent plasma transfusion for the treatment of COVID-19: Systematic review. *Journal of Medical Virology*. 2020;92(9):1475–1483. Available from: https://dx.doi.org/10.1002/jmv.25961.

21) Ahmad S. A Review of COVID-19 (Coronavirus Disease-2019) Diagnosis, Treatments and Prevention. *Eurasian J Med Oncol*. 2020;4(2):116–125. Available from: https://doi.org/10.14744/ejmo.2020.90853.
22) Krishnakumar B, Rana S. COVID 19 in INDIA: Strategies to combat from combination threat of life and livelihood. *Journal of Microbiology, Immunology and Infection*. 2020;53(3):389–391. Available from: https://dx.doi.org/10.1016/j.jmii.2020.03.024.
23) Gupta R, Pal SK, Pandey G. 2020.
24) Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 Vaccines at Pandemic Speed. *New England Journal of Medicine*. 2020;382(21):1969–1973. Available from: https://dx.doi.org/10.1056/nejmp2005630.
25) Fong SJ, Li G, Dey N, Gonzalez-Crespo R, Herrera-Viedma E. Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak. *International Journal of Interactive Multimedia and Artificial Intelligence*. 2020;6:132–132. Available from: https://dx.doi.org/10.9781/ijimai.2020.02.002.
26) Tandon H, Ranjan P, Chakraborty T, Suhag V. Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. 2020.
27) Todelo TL, Aliac CJG. Predictability of Earthquake Occurrence Using Auto Regressive Integrated Moving Average (ARIMA) Model. 2019.
28) Gupta R, Pal SK. Trend Analysis and Forecasting of COVID-19 outbreak in India. 2020. Available from: https://doi.org/10.1101/2020.03.26.20044511.
29) Kumar P, et al. Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020 : ARIMA Model with Machine Learning Approach. 2020.
30) He Z, Tao H. International Journal of Infectious Diseases Epidemiology and ARIMA model of positive-rate of in fl uenza viruses among children in Wuhan , China : A nine-year retrospective study. *Int J Infect Dis*. 2018;74:61–70. Available from: https://doi.org/10.1016/j.ijid.2018.07.003.
31) Shinde GR, Kalamkar AB, Mahalle PN. Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art. 2020.
32) Shinde RG, Kalamkar BA, Mahalle NP, Dey N, Chaki J, Hassanien AE. Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. *SN Computer Science*. 2020;1(4):1–15. Available from: https://dx.doi.org/10.1007/s42979-020-00209-9.
33) Song X. Time series analysis of in fl uenza incidence in Chinese provinces from 2004 to 2011. 2004.
34) Zeroual A, Harrou F, Dairi A, Sun Y. Chaos , Solitons and Fractals Deep learning methods for forecasting COVID-19 time-Series data : A Comparative study. 2020;140. Available from: https://dx.doi.org/10.1016/j.chaos.2020.110121.
35) Salgotra R, Gandomi M, Gandomi AH. Chaos , Solitons and Fractals Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming. 2020. Available from: https://doi.org/10.1016/j.chaos.2020.109945.
36) Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*. 2020;29. Available from: https://dx.doi.org/10.1016/j.dib.2020.105340.
37) Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*. 2020;1:12–18. Available from: https://dx.doi.org/10.1016/j.jnlssr.2020.06.007.
38) Gola A, Kumar R. Review of Forecasting Models for Coronavirus (COVID-19) Pandemic in India during Country-wise Lockdown. 2020.
39) Samanta A. Real-time Forecasting of COVID-19 Prevalence in India using ARIMA Model. 2020. Available from: https://doi.org/10.35940/ijmh.J0973.0641020.
40) Simon JC, Fong J, Dey N. Artificial Intelligence for Coronavirus Outbreak. Singapore. Springer. .
41) Nau R. Mathematical structure of ARIMA models. 2014.
42) Chauhan B, Kumar S, Tripathi A, Malik RP. Modified 2D Proca Theory: Revisited under BRST and (Anti-)chiral Superfield Formalisms. *Advances in High Energy Physics*. 2020;2020:1–38. Available from: https://dx.doi.org/10.1155/2020/3495168.